

A cross-linguistic study of diphthongs in spoken word processing in Japanese and English*

Kiyoko Yoneyama
yoneyama@ling.ohio-state.edu

Abstract: This paper investigates the proper treatment of diphthongs in Japanese and English in terms of spoken word processing. Three phoneme-monitoring experiments were conducted with three different groups of language users: Japanese monolinguals, English monolinguals and semi-bilingual Japanese speakers of English; both English and Japanese materials were used. The results showed that English monolinguals treat diphthongs as single units during language processing, while Japanese monolinguals treat them as two separate units. The processing of Japanese and English diphthongs by semi-bilingual Japanese speakers of English is also discussed.

INTRODUCTION

Spoken words have a rich structural organization in memory consisting of both syllabic and subsyllabic representations. Previous cross-linguistic investigation into segmenting speech has shown that word processing in different languages involves a variety of linguistic units. Results from French, for example, suggest that listeners in that language segment speech at syllable boundaries (Mehler et al., 1981). English listeners, in contrast, segment speech at the onset of strong (but not weak) syllables (Cutler and Norris, 1988). Furthermore, Japanese listeners segment speech at mora boundaries (Otake et al., 1993; Cutler and Otake 1994; Otake et al., 1996; Yoneyama, 1995). Collectively, these strategies are referred to as "rhythm-based segmentation strategies (RSS)".

In addition to an RSS, auditory recognition models such as the Cohort model (Marslen-Wilson & Welsh, 1987), TRACE (McClelland & Elman, 1986), and SHORTLIST (Norris, 1994), all assume that segmentation of speech into words is achieved as an automatic consequence of lexical access from the sub-lexical phoneme- or feature-recognition process. Therefore, adult listeners with mature lexicons can map speech onto words directly without reference to language-specific prosodic units. This type of segmentation is referred to as the "general segmentation strategy (GSS)", and is available for all language users.

The focus of the present study is Japanese listeners' sensitivity to moraic structure. The Japanese mora is a component of syllable structure. Light syllables consist of one mora while syllables with a complex vowel or a coda consist of two. Japanese has a strict phonological structure and there are only five types of mora: CV, CCV, V, nasal coda (represented as N) and geminate consonant (represented as Q).

* This study is supported by Department of Linguistics and Department of Psychology. I would like to express my sincere gratitude for valuable comments and guidance provided by Mary Beckman, Jose Benki, Anne Cutler, Keith Johnson, James McQueen, and Takashi Otake. Thanks also to Mark Pitt, who enabled me to test the English listeners, and Katherine Smith and Lisa Shoaf for their technical support. I also extend my gratitude to editors, Kim Ainsworth-Darnell and Mariapaola D'Imperio for their comments on my earlier version of this paper. Of course, all errors are mine.

Japanese listeners appear to be able to exploit the moraic structure of words in a variety of different ways. Otake et al. (1993) presented Japanese listeners with sequences of natural spoken words and asked them to indicate when they detected a word beginning with a particular CV sequence, for example, /ta/. Listeners were equally fast to identify the /ta/ target in *tanishi* and in *tanshi* because both begin with the same mora /ta/. However, when the target was /tan/, Japanese listeners had difficulty detecting the target in *tanishi* because it corresponds to the whole first mora and a part of the second. Also, Cutler and Otake (1994) found that Japanese listeners detected phoneme targets which were moras in themselves more rapidly than targets which formed only part of a CV mora. Thus, for example, /n/ was detected more rapidly in *inka* than in *inori*. /o/ was detected more rapidly in *aokabi* than in *kokage*.

The vowel target /o/ in Cutler and Otake (1994) was a moraic vowel, or one that was a syllable by itself. However, a moraic vowel can occur tautosyllabically, following another vowel. Such vowel-vowel sequences are often called diphthongs. According to Vance (1987), /ai/, /ei/, /oi/, /ui/, /au/ and /ou/ can be considered diphthongs in Japanese, but only if they occur in a single morpheme and if the second segment does not bear an accent. Assuming that the previous studies concerning Japanese listeners' sensitivity to mora are correct, these people should have no problem detecting the second part of a diphthong because it is considered as one mora. Experiment 1 tests this hypothesis. Hypothesis is a comparison: If Japanese listeners process speech mora by mora in a mora-based RSS, then they should identify [i] in a diphthong (a V mora) more quickly and accurately than [i] in a CV mora.

EXPERIMENT 1

PARTICIPANTS

Twenty-four Dokkyo University undergraduates participated in exchange for course credit. All were native speakers of Japanese and reported no hearing difficulties. All had studied English only in school and none had ever stayed in an English-speaking country for more than 3 months.

MATERIALS AND PROCEDURE

Two sets of language materials (Japanese and English) were constructed. The Japanese materials, shown in Table 1, consist of 24 content words (nouns, verbs, and adjectives). Half contained [oi] (CV₁V₂ words) and the other half contained [i] (CV₁ words). They formed twelve pairs, contrasting in the occurrence of [o] in the first syllable. Each of CV₁V₂ words contained a diphthong in Japanese, which follows Vance's (1987) definitions. We admit that because a number of words which contain an [oi] sequence is limited in Japanese, most of the stimuli are borrowed words¹.

CV ₁ V ₂ words			CV ₁ words		
koin	boisu	join	rin	nisu	misu
boiru	doitsu	roido	misa	piza	bika
koiru	hoiru	hoiro	pin	kika	piru
noizu	koika	toire	bisu	biru	kigo

Table 1: Japanese stimulus words

¹ Of CV₁V₂ words, only "koika" is a Sino-Japanese word, and the others are borrowed words.

The English materials, shown in Table 2, are very similar in the Japanese materials. As before, there are 24 content words (nouns, verbs, and adjectives). Half contained [oi] (CV₁V₂ words) and the other half contained [i] (CV₁ words). They formed twelve pairs, contrasting in the occurrence of [o] in the first syllable.

CV ₁ V ₂ words			CV ₁ words		
join	boil	moist	pit	kin	mist
foil	coil	soil	gin	bill	win
toil	voice	loin	fill	kill	sit
coin	void	joint	till	mill	fit

Table 2: English stimulus words

The vowel sequences [oi] and [oi] were chosen for my stimuli because they are represented by two letters in both English and Japanese. Thus, any orthographic influence on responses should be equal across languages.

The English materials were recorded by a male native speaker of American English (Northern Kentucky dialect), while the Japanese materials were recorded by a male native speaker of Japanese (Tokyo dialect). All materials were stored on DAT tapes and spoken at a normal rate of speech.

The target words were mixed with filler words and were arranged into 48 word sequences in both sets of materials. In each set of materials, half contained one of the experimental target words and the other half did not. Each word sequence varied from two to six words. Of the twenty-four sequences which did not contain one of the experimental target words, the half contained a dummy target, which elicited participants' responses. A target always occurred in the penultimate position in each word sequence.

Stimuli were presented to participants binaurally over headphones at a comfortable listening level. Responses were collected by pressing a button on a response board. Stimulus presentation and data collection were controlled by a PC computer. Participants were tested in a quiet room. The English and Japanese materials were presented in separate session, with the English session coming first. In each test session, participants listened to 48 word sequences. For each sequence, they were instructed to think constantly of the target sound and to respond as soon as they detected it. The target sound in each test session was different. In the English test session, it was a high front lax vowel ([ɪ]) and in the Japanese test session, it was a high front tense vowel ([i]). Participants were not informed about the experimental manipulations or about the characteristics of the stimuli. A 10-sequence practice session preceded the 48 test sequences, which were presented in a fixed order in each test session. The entire experiment lasted approximately 50 minutes.

RESULTS

Analyses of variance were conducted separately on the mean rates of missed data and mean response times. Miss rate is a percentage of trials on which the listeners failed to detect the target. RT rate is time in ms between the target sound onset and the listener's detection response. The condition means of miss rate are give in Figure 1. The analysis of miss rate showed that Japanese targets were detected significantly more accurately (19.3%) than English ones (87.6%; $F[1, 1148] = 1043.439, p < .0001$). Also, targets in CV₁V₂ words were detected more accurately (50%) than those in CV₁ words (57.2%; $F[1, 1148] = 11.242, p < .001$). There was an interaction between material language and word type: targets in CV₁ words in the Japanese materials were detected more accurately than those in the English materials ($F[1, 1148] = 4.875, p < .05$). In English materials, targets in CV₁V₂ words were detected more accurately than those in CV₁ words ($F[1, 576] = 19.366, p < .0001$). However, the same effect was not observed in the Japanese materials.

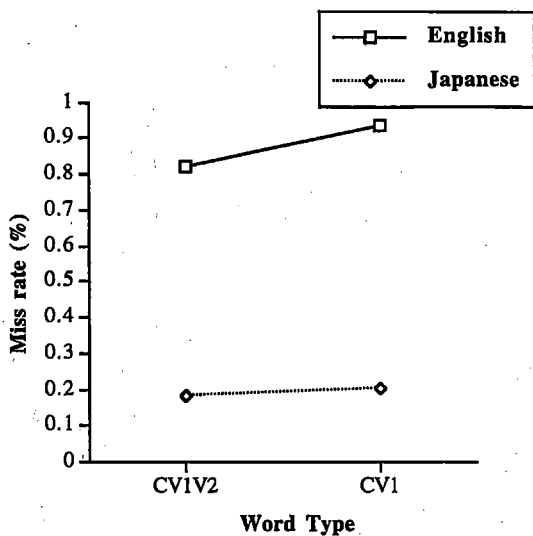


Fig.1: The mean miss rate in (%) as a function of stimulus word type and material language, Experiment 1.

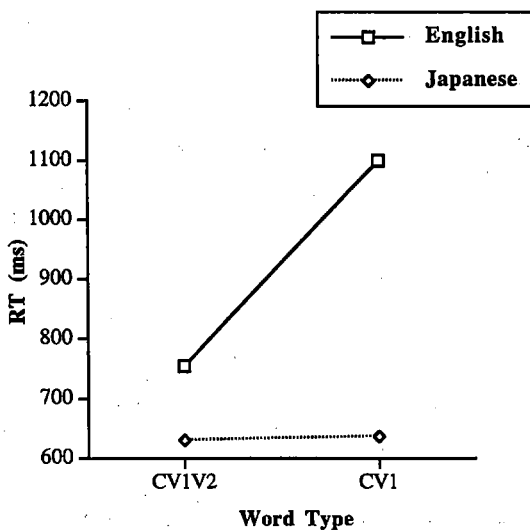


Fig.2: The mean response times in ms. as a function of stimulus word type and material language, Experiment 1.

The condition means of response times are given in Figure 2. The analysis of response times showed that Japanese targets were detected significantly faster (633 ms) than English ones (844 ms; $F[1, 531] = 1.07, p < .0001$). It also showed that targets in CV₁V₂ words were detected significantly faster (654 ms) than those in CV₁ words (668 ms; $F[1, 531] = 37.11, p < .0001$). There was an interaction between material language and word type: in the English materials, targets in CV₁V₂ words were detected significantly faster (756 ms) than those in CV₁ words (307 ms), and in CV₁ words, Japanese targets were detected significantly faster (631 ms) than English ones (1100 ms; $F[1, 531] = 35.565, p < .0001$).

DISCUSSION

Experiment 1 provides contrasting evidence concerning whether or not Japanese listeners use general and rhythm-based segmentation strategies. There was no evidence of a moraic effect with the Japanese materials: responses to the targets in CV₁V₂ words were as fast and accurate as those in CV₁ words. If the participants had employed moraic segmentation, they should have been able to detect targets in CV₁V₂ words significantly more quickly and accurately than those in CV₁ words. The responses to English materials, conversely, suggest that Japanese listeners are sensitive to the moraic structure, which is consistent with Cutler and Otake (1994). Both the miss rate analysis and the response time analysis showed a moraic effect: targets in CV₁V₂ words were detected significantly faster and more accurately than those in CV₁ words.

One possible explanation for these results is that the participants employed a GSS when they listening to Japanese and an RSS when listening to English. Cutler et al. (1992) have hypothesized that two segmentation strategies are available for processing of a native language. Young children use an RSS to construct their lexicon. As their lexicon matures, however, they learn to segment speech directly into words, without any phonological units intervening (GSS). Since all participants in this experiment have a mature Japanese lexicon, it is not surprising that they may have employed a GSS in the processing of Japanese, their native language. This would be consistent with the findings in Yoneyama (1995), where Japanese listeners showed the same GSS as English and French listeners in Cutler et al. (1986).

My explanation for the responses to English materials is consistent with Cutler and Otake (1994). Cutler and Otake reported that their Japanese participants are sensitive to moraic nasals even in English, as if they were listening to Japanese. This could indicate that Japanese listeners employ an RSS when they are listening to a non-native language. My Japanese participants showed moraic sensitivity to the vowels in English diphthongs.

One concern regarding my conclusion here is why the Japanese participants did not show their sensitivity to moraic structure in Japanese diphthongs, even though all previous on-line experiments have found such effect. One possible explanation is that Japanese listeners are also sensitive to syllable structure. Notice that Cutler and Otake (1994) showed a clear moraic effect using vowel targets in Japanese materials. On the other hand, the current study did not show this effect even though targets were second half of the vowel sequence. The only difference between two studies was position of targets in terms of syllable structure. Each of Japanese vowel targets of Cutler and Otake (1994) was also considered as one syllable by itself. Conversely, each of my Japanese vowel targets was a part of nucleus and it did not become a syllable by itself. Another possibility is that Japanese listeners treat diphthongs as single units even when they listening to Japanese. Thus, since the bond between two vowels in diphthongs is so tight, the second vowel of a diphthong was hard to be detected even though it had a moraic status. In any case, this inconsistency is certainly a topic for future study.

In Experiment 2, I will examine how English listeners treat diphthongs in English and Japanese. Both Otake et al. (1993) and Cutler and Otake (1994) found that English speakers did not show any sensitivity to moraic structure. Cutler et al. (1992) hypothesized that the English participants in these studies employed an RSS when listening to Japanese. If this is true, English listeners in the current study should also use the same

listening strategy with English and Japanese materials, and show no sensitivity to moraic structure.

EXPERIMENT 2

PARTICIPANTS

Nineteen Ohio State University undergraduates participated in exchange for course credit. All were native speakers of English and reported no hearing difficulties. None have ever either studied Japanese or stayed in Japan for more than 3 months.

MATERIALS AND PROCEDURE

The materials were the same as in Experiment 1. However, in this experiment, all the English and Japanese stimulus words were digitally recorded onto computer disk (sampling rate of 10 kHz, low-pass filtered at 4.8 kHz) and edited and saved as separated sound files. Stimuli were presented to participants binaurally over headphones at a comfortable listening level. Responses were collected by pressing the button on a response board. Stimulus presentation and data collection were controlled by a PC/AT computer.

The procedure was generally the same as in Experiment 1, with a few minor changes. As many as four participants at a time were tested simultaneously in individual sound-attenuated booths. Also, the order of two test sessions was counterbalanced; half of the participants were presented the English test session first and the latter half were presented the Japanese test session first. The entire experiment lasted approximately 50 minutes.

RESULTS

Analyses of variance were conducted separately on the mean rates of missed data in each condition and the mean response time in each condition, separately. The condition means of miss rate are give in Figure 3. The analysis of miss rate showed that the target in CV₁ words was detected more accurately (36.7%) than that in CV₁V₂ words (72.2%; $F [1, 908] = 133.281, p < .0001$). The results showed an interaction between material language and word type: the target in CV₁ words in the Japanese materials was detected less accurately than that in the English materials and the target in CV₁V₂ words in the English materials was detected less accurately than that in the Japanese materials ($F [1, 908] = 6.582, p < .0105$). No effect of material language was observed. Further analyses showed that in both language materials, CV₁ words were detected more accurately than CV₁V₂ words ($F [1, 454] = 106.599, p < .0001$ for English; $F [1, 454] = 37.813, p < .0001$ for Japanese).

The condition means of response times are give in Figure 4. The analysis of response times showed that Japanese stimulus words were detected significantly faster (715 ms) than English ones (788 ms; $F [1, 386] = 25.17, p < .0001$). In addition, targets in CV₁ words were detected significantly faster (725 ms) than that CV₁V₂ words (806 ms; $F [1, 386] = 19.996, p < .0001$). There was an interaction between material language and word type: in both languages, targets in CV₁ words were detected significantly faster than in CV₁V₂ words ($F [1, 386] = 13.301, p < .0003$). However, the response time difference between two conditions in English was much larger (206 ms) than that in Japanese materials (21 ms).

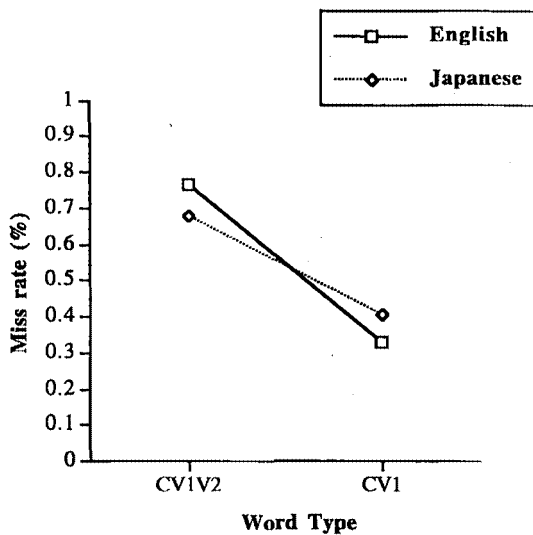


Fig.3: The mean miss rate in (%) as a function of stimulus word type and material language, Experiment 2.

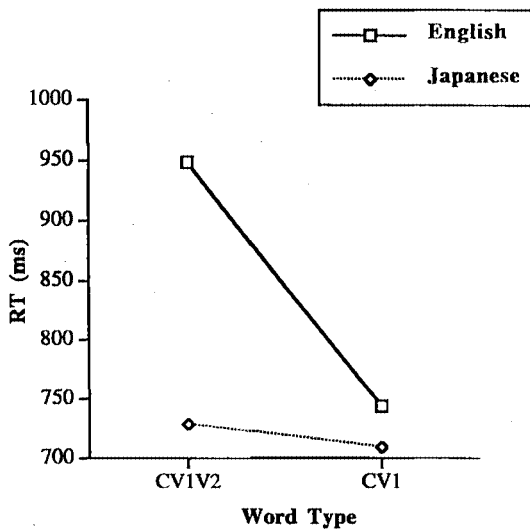


Fig.4: The mean response times in ms. as a function of stimulus word type and material language, Experiment 2.

DISCUSSION

The miss rate analysis showed that English listeners treated diphthongs in English and Japanese in the same way. They consistently responded to targets in CV₁ words more accurately than to CV₁V₂ words. The analysis of response time showed the same effect. If they were sensitive to moraic structure, the results should be opposite: targets in CV₁V₂ words were detected faster and more accurately than those in CV₁ words. This not being the case, it appears that my English participants were not sensitive to moraic structure.

My results here confirmed the previous on-line studies with English listeners when listening to Japanese. Both Otake et al. (1993) and Cutler and Otake (1994) reported that English speakers seemed to process Japanese as if they were listening to English. It seems feasible that the English listeners in my study may have employed a native way of listening even when listening to a non-native language. In other words, my English participants might have employed their English RSS when listening to Japanese.

One concern regarding this conclusion is that my experiment here cannot directly show the English speakers' stress-timed segmentation strategy, which was originally confirmed by Cutler and Norris (1988). These researchers found that English listeners segment speech at the strong syllables, which needs the vertical specification of speech (strong versus weak stress assignment). However, I assume that the effect of material language in reaction time analysis might show some language-specific way of listening by my English participants.

Notice that my English participants listened to two set of language materials, both English and Japanese language materials, and that in each set, the target sound was different. In the English materials, the target sound was always [i] whereas in the Japanese materials, it was always [i]. Thus, the effect of material language might be rephrased as the effect of the target sound difference in two languages. The reaction time analysis might indicate that [i] can be detected faster than [i] across languages. Findings in van Ooijen (1994) support this idea. She investigated the processing of vowels and consonants by English listeners, and conducted 9 reaction-time experiments. Using her list of individual target bearing items together with their mean reaction time in (ms) across participants, it is possible to estimate the processing time of these two vowels. Table 3 shows the from van Ooijen's (1994) Experiments 5 and 9, that have the same structure of my English materials. All the words have the primary stress on the first syllable. An analysis of variance I calculated from these words showed that /i/ was detected significantly faster (414 ms) than /u/ (531.5 ms; $F[1, 12] = 23.06, p < .004$), which is consistent with my finding.

words with /i/		words with /u/	
feast	458	tissue	494
sheep	483	mitten	581
priest	439	liver	514
seek	354	wither	502
leave	358	sickle	525
scream	397	mistress	573
needle	373		
feeling	454		
Mean	414.5	Mean	531.5

Table 3: words partially from Experiments 5 and 9 in van Ooijen (1994) that contain either /i/ or /u/.

In any case, we can be sure that acoustic information can affect participants' performance when processing spoken materials in one's native and non-native language. This suggests that understanding the mechanism between the segmentation and acoustic information is crucial for revealing spoken word processing.

Experiment 3 investigates how semi-bilingual Japanese speakers of English treat English and Japanese diphthongs in spoken word processing. Of interest is whether or not knowledge of a second language can influence processing in the first, or vice versa. Participants in this experiment were native speakers of Japanese who has lived in an English speaking country for more than 3 years (semi-bilinguals). Because Japanese is their first language, I expect that they will treat Japanese diphthongs like their monolingual Japanese counterparts. But how will these individuals treat English diphthongs? Cutler et al. (1992) have hypothesized that bilinguals and monolinguals perform differently when they listen to a non-native language. This might indicate that diphthongs in English words stored in the lexicon of semi-bilinguals might be treated differently from those for Japanese words, depending on the degrees of foreign-language exposure. The last experiment explores this possibility.

EXPERIMENT 3

PARTICIPANTS

Twenty-two Dokkyo University undergraduates participated in exchange for a nominal fee. All had good communicative abilities in English, and had lived for a minimum of three years in an English-speaking country, but were recognizably not native speakers of English. They reported no hearing difficulties.

MATERIALS AND PROCEDURE

Materials and procedure were the same as in Experiment 1.

RESULTS

Analyses of variance were conducted separately on the mean rates of missed data in each condition and the mean response time in each condition. The condition means of miss rate are give in Figure 5. The analysis of miss rate showed an interaction between material language and word type: targets in CV_1V_2 words were detected significantly more accurately (14%) than those in CV_1 words (18.9%) in Japanese, and CV_1 words were detected significantly more accurately (12.1%) than CV_1V_2 words in English (19.3%; $F[1, 1052] = 7.204, p < .01$). No other main effect was observed.

The condition means of response times are give in Fig. 6. No main effect or an interaction were observed.

DISCUSSION

This experiment with semi-bilinguals confirmed the previous findings with Japanese monolingual and bilingual listeners. First, the analysis of miss for Japanese materials showed that my semi-bilingual listeners are sensitive to moraic structure when listening to Japanese. They responded to targets in CV_1V_2 words more accurately than to those in CV_1 words, suggesting that they employed a Japanese RSS. This finding is consistent with previous on-line studies with Japanese listeners (Otake et al., 1993; Cutler et al. (1994). Another finding from the miss rate analysis is that my semi-bilinguals might have employed a GSS when listening to English, which is consistent with the findings by semi-bilinguals Japanese speakers of English in Experiment 1 in Yoneyama (1996).

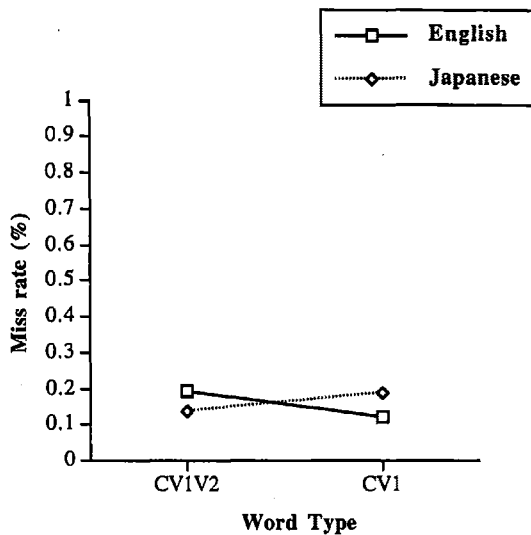


Fig.5: The mean miss rate (%) as a function of stimulus word type and material language, Experiment 3

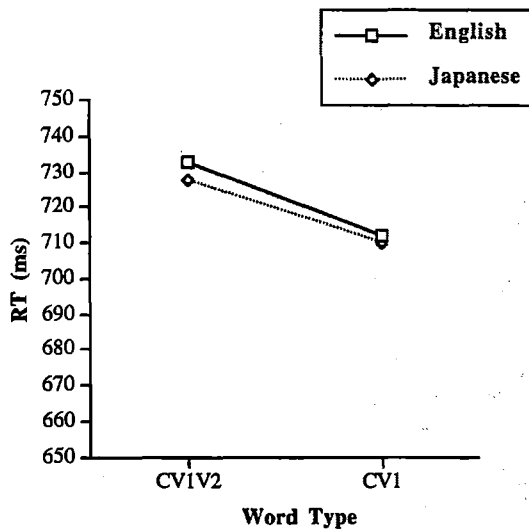


Fig.6: The mean response times in ms. as a function of stimulus word type and material language, Experiment 3.

These findings might indicate that semi-bilinguals can choose either a GSS and an RSS, depending on the language input. One possible explanation is that the second language exposure might enable them to shift the application of their RSS (moraic segmentation) which do not work well with English, to the one of a GSS. This is exactly what Cutler et al. (1992) have claimed: bilingual speakers have an ability to suppress their native segmentation strategy to a non-native language.

Of course, it is possible that the semi-bilinguals in the current study may have been sufficiently fluent in both the first and second languages to rely on segmental information no matter which segmentation strategy they are using. The miss rate in the four conditions were very low and the overall miss rate was 15.5%. Surprisingly, even when we compare them with English listeners in CV₁ word condition in English materials, performance of the semi-bilinguals was better (12.1%) than that of English listeners (32.9%) from Experiment 2. Also, the results of response time analysis showed neither significant main effects nor an interaction. These findings might suggest that the semi-bilingual participants may be able to use segmental information equally in English and in Japanese.

GENERAL DISCUSSION

The current study further investigates the sensitivity to moraic structure of three groups of language users: Japanese listeners, English listeners and semi-bilinguals Japanese speakers of English. Experiment 1 was designed to further investigate the sensitivity to moraic structure by Japanese listeners, using two sets of language materials (English and Japanese). The Japanese listeners showed a moraic effect in English materials, although they did not show it in Japanese materials, where it has been found in the previous studies. Experiment 2 investigated how English listeners responded to the same materials used in Experiment 1. The results showed that the English listeners were not sensitive to moraic structure either in English or Japanese, and seemed to listen to both English and Japanese in the same way. Experiment 3 tested how semi-bilingual Japanese speakers of English responded to the same materials used in Experiment 1. The results indicated that the semi-bilinguals showed a moraic effect in Japanese materials while they did not in English materials.

The possible comparisons among these three different groups of language users shed light on further aspects of segmenting speech. First of all, from the results in Experiments 1 and 3, we see native speakers of Japanese are generally sensitive to moraic structure. The semi-bilinguals showed their sensitivity in Japanese materials, which clearly supports Otake et al. (1993) and Cutler and Otake (1994).

One concern with our conclusion is that my Japanese listeners in Experiment 1 did not show a clear moraic effect in Japanese materials, which was supposed to show from the findings in the previous studies. Even though we have not yet figured out why my Japanese listeners employed a unexpected GSS to process Japanese diphthongs, I assume that some unknown factor, such as syllable structure and acoustic property of diphthongs affected my participants' performance, and made them switch their listening strategy from an RSS to a GSS. Since my Japanese listeners showed their sensitivity even to English materials, which indicates a typical characteristic of responses by monolingual listeners, I hypothesize that the Japanese participants in Experiment 1 are sensitive to moraic structure, even though the results did not show it as clearly as was hoped. Both Experiments 1 and 3 showed the sensitivity to moraic structure by native speakers of Japanese. They confirmed the previous on-line studies with Japanese listeners.

A comparison between native and non-native speakers of Japanese will allow us to explore the language-specific aspects of listening. As we have shown above, native speakers of Japanese showed the sensitivity to moraic structure. Conversely, Experiment 2 showed, that English listeners did not show the same sensitivity, as in the previous studies. These results confirmed that only Japanese natives are sensitive to moraic structure.

In addition to the sensitivity to moraic structure by Japanese natives, the data of semi-bilinguals in this study suggests that extensive second-language experience enables people to treat diphthongs differently in different languages. As Cutler et al. (1992) have

proposed, the main difference between bilingual and monolingual speakers is that the former have a lexicon of a familiar non-native language whereas the latter does not. This means that semi-bilingual Japanese speakers must have acquired English words in their lexicon as well, even though English is their second-language. Their extensive language experience enables bilingual speakers to suppress their native rhythm-based strategy when dealing with foreign language input. Experiment 3 showed that semi-bilinguals treat diphthongs differently, depending on which language they are listening to.

Finally, I would like to mention about the relation between diphthongs and language processing. Japanese listeners in the current study did not have difficulty detecting the second vowel of diphthongs in English and Japanese. This suggests that they treat diphthongs as two separate units. On the other hand, English listeners in Experiment 2 showed a difficulty detecting the same targets in the same materials. This might indicate that they treat both Japanese and English diphthongs as single units in the language processing. Interestingly, the semi-bilinguals in Experiment 3 seem to treat English and Japanese diphthongs differently. In Japanese, they seem to employ a moraic segmentation, and to treat Japanese diphthongs as two separate units, just like monolingual Japanese listeners in Experiment 1. On the other hand, when they listen to English, they seem to treat English diphthongs differently. In English materials, they responded to targets in CV₁ word more accurately than to those in CV₁V₂ words. This might indicate that they employ a general segmentation strategy. Also a clear interaction between language material and a word type might suggest that they also treated English diphthongs as single units, like English monolinguals in Experiment 2. However, as we have seen in Experiments 2 and 3, the semi-bilinguals listening to English detected a high front lax vowel in CV₁V₂ words in 80% of the trials, even though English listeners detected the same target in the same materials in only 20% of the trials. They were definitely able to identify the target sound in English diphthongs whereas English monolinguals were not. This might also suggest that performance by semi-bilinguals when processing in English is different from that by English monolinguals. Together, the results in Experiment 3 might be interpreted as follows. Firstly, second language experience may enable individuals to treat diphthongs differently in that language. This language knowledge enabled them to switch from an RSS to a GSS, which confirmed the previous bilingual studies (Cutler et al., 1992; Bradley et al., 1993; Kearns 1994; Yoneyama 1996). Secondly, however, they cannot fully suppress their native way of listening. The results showed that they did not have difficulty accessing to moraic structure even in English. This is because they are not able to fully suppress their native RSS when they processing in English. My semi-bilinguals are not "perfect" bilinguals as those in Cutler et al. (1992). This finding is coincide with the findings with semi-bilingual Japanese speakers of English in Yoneyama (1996).

In this paper, I claim that, in terms of spoken language processing, Japanese diphthongs should be treated as two separate units in reference to moraic structure, and English diphthongs should be treated as single units. However, I do not attempt to represent Japanese and English diphthongs in syllable structures here. This should be investigated in further studies.

REFERENCES

- Bradley, D.C., Sanchez-Casas, R.M., & Garcia-Albea, J.E. (1993). "The status of the syllable in the perception of English and Spanish," *Language and Cognitive Processes*, 8, 197-233.
- Cutler, A., Mehler, J., Norris, D. G. & Segui, J. (1992). "The monolingual nature of speech segmentation by bilingual," *Cognitive Psychology*, 24, 381-410.
- Cutler, A. & Norris, D. G. (1988). "The role of strong syllables in segmentation for lexical access," *Journal of Experimental Psychology: Human Perception & Performance*, 14, 113-121.
- Cutler, A. & Otake, T. (1994). "Mora or Phoneme? Further evidence for language-specific listening," *Journal of Memory & Language*, 33, 824-844.

- Kearns, R.K. (1994). *Prelexical speech processing in mono- & bilinguals*, Ph.D. Dissertation, University of Cambridge.
- Marslen-Wilson, W.D. & Welsh, A. (1978). "Processing interaction and lexical access during word recognition in continuous speech," *Cognitive Psychology*, 10, 29-63.
- McClelland, J.L. & Elman, J.F. (1986). "The TRACE model of speech perception," *Cognitive Psychology*, 18, 1-86.
- Mehler, J., Dommergues, J.-Y., Frauenfelder, U. & Segui, J. (1981). "The syllable's role in speech segmentation," *Journal of Verbal Learning & Verbal Behavior*, 20, 298-305.
- Norris, D.G. (1994). "SHORTLIST: A hybrid connectionist model of continuous speech recognition," *Cognition*, 52, 189-234.
- Otake, T., Hatano, G., Cutler, A. & Mehler, J. (1993). "Mora or syllable? Speech segmentation in Japanese," *Journal of Memory & Language*, 32, 258-278.
- Otake, T., Hatano, G., & Yoneyama, K. (1996). "Japanese speech segmentation by Japanese listeners," in T. Otake & A. Cutler eds., *Phonological structure and language processing: Cross-linguistic studies*, 183-201, Mouton de Gruyter, Berlin.
- van Ooijen, B. (1994). *The processing of vowel and consonant*, Ph.D. dissertation, University of Leiden, The Netherlands.
- Vance, T. J. (1987). *An Introduction to Japanese Phonology*. State University of New York Press.
- Yoneyama, K. (1995). *Segmentation procedure by semi-bilingual speakers of Japanese and English*. An unpublished M.A. thesis, Dokkyo University, Soka, Japan.
- Yoneyama, K. (1996). "Segmentation strategies for spoken language recognition: Evidence from Semi-bilingual Japanese speakers of English," *Proceedings of the 1996 International Conference on Spoken Language Processing*, Philadelphia, vol. 1, 454-457.